关于新型冠状病毒感染人数的预测研究

吴宇宸[1,2]
1：中国科学院大学，北京 100049
2：中国科学院院数学与系统科学研究院，北京 100190
wuyuchen@lsec.cc.ac.cn
2020 年 2 月 16 日

[摘要] 2019 年 12 月，新型冠状病毒肺炎(NCP，又称 2019-nCoV)疫情从武汉开始爆发,几天内迅速传播到全国乃至海外，对我国的工农业生产和人民生活产生了重要影响。科学有效掌控疫情发展对疫情防控至关重要。本文基于中国卫健委及湖北省卫健委每日公布的累计确诊数，采用逻辑斯蒂模型对数据进行了拟合，以期给该疾病的防控治提供科学依据。通过公布的疫情数据，我们反演了模型的参数，进而有效地模拟了目前疫情的发展，并预测了疫情未来的趋势。我们预测，湖北省疫情还要持续至少 2 周，而在全国其他地区，疫情可望 1 周左右达到顶峰。

[关键词] 冠状病毒感染肺炎 数学预测 湖北省 全国其它地区
MR（2000）主题分类：93A30

MATHEMATICAL FORECASTING OF THE NOVEL CORONAVIRUS PNEUMONIA (NCP)

Wu Yuchen1,2

1：University of Chinese Academy of Sciences, Beijing 100049

2：Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190

wuyuchen@lsec.cc.ac.cn

Abstract

Novel Coronavirus Pneumonia (NCP, or alternatively 2019-nCoV), initially blown up in Wuhan in December of 2019, has been quickly spread all over China, and even other countries of the world, which has produced an important effect to the agricultural and industrial activities, and daily life. It is expected that a well-known recognition is essential to the effective prevention of the disease. Based on the daily announced numbers of the infective

people from the National and Hubei provincial Health commissions, a logistic model is applied in this paper for data fitting, in order to provide some scientific information for the effective prevention and controlling of the disease. Using the parameters obtained from the data simulation, a susceptible-infected (SI) model is used to forecast the future trend of the NCP. Our work indicated that the epidemic will last at least two additional weeks in Hubei, but should come to an apex in one week in other areas of China.

Keywords: Novel Coronavirus Pneumonia (NCP); Mathematical forecasting; Hubei Province, Other areas of China

2000 Mathematics Subject Classification: 93A30

1 引言

2019 年底至 2020 年初，我国武汉爆发新型冠状病毒感染的肺炎，引起全国人民的关注。自 2020 年 1 月 23 日武汉封城至今已经三周，该传染性疾病的发展趋势对未来如何部署新的防控治策略，乃至国内生产的恢复都有着非常重要的影响。鉴于这些与人民生活密

切相关的重要问题，国内外学者运用不同的数学模型，对该疾病的扩散与传播做了一定的工作[1-3]，但就该疾病未来的发展趋势，目前开展的研究并不多见。因此，本文利用目前被大多数学者认可的数学模型，对冠状病毒感染肺炎疾病的发展趋势进行预测，以期对目前的防控治工作提供科学参考。

2 模型建立

设 y 为已感染人数，t 为时间，在总人数趋于无穷的情况下（或者说，在已感染人数相对总人数是一个可以忽略的量的情况下），此时，感染人数增加的速率与已感染人数成正比，因此我们可以得到常微分方程 y’=ay。解这个微分方程，我们可以得到，等比数列上升的趋势图，这也是疫情初期感染人数差不多是等比上升的原因。

不过，总人数其实是有限的。尤其是在做了隔离措施后，一个人能接触的人也是有限的。这样，感染人数的增加实际上与已感染人数和未感染人数乘积成正比。再加上有了隔离措施，切断了大多数人与病人的联系，因此我们可以列出如下常微分方程 y’=ay(b-y)。其中，a 代表传染率，b 代表可能感染的总人数。这个方程的解要稍微复杂一些，可以通过以下步骤解出：

那么最终，我们可以得到

实际上，这个模型就是传染病的 SI 模型[4]。其微分方程的解有解析表达式，易于优化求解，但没有考虑病人的康复，以及潜伏期的情况。若考虑康复，则得到 SIR 模型[5]，若再考虑潜伏期，即得到 SEIR 模型，甚至更复杂的 SEIJR 模型[6]。尽管传染病研究多采用 SIR 或 SEIR 模型，但这两个模型的解均没有显式表达，这意味着需要用数值方法解常微分方程，这意味着后面的数据拟合，也就是优化过程将会非常复杂，且在疫情初期康复人数相比感染人数可以忽略不计。因此，我们仍采用相对简单的 SI 模型，以期得到较为可靠的模拟结果。

这些参数中，A 代表传染的速率，K 代表最终传染病人累计的数量，K/P 大概是第一天新型冠状病毒肺炎的病例数量。

3 数据拟合

上述讨论只是一个理论分析，并没有结合本次疫情的数据。而且，其中的 k,b,C,a 等变量都极难通过经验确定，因此我们要结合我们已知的数据，确定那些变量以及 K,P,A 的值，这个过程就是数据拟合的过程。

表 1 冠状病毒肺炎疫情

日期	全国确诊	全国新增	湖北确诊	湖北新增	其它地区确诊	其它地区新增
1 月 19 日	214	-	198	-	16	-
1 月 20 日	291	77	270	72	21	5
1 月 21 日	440	149	375	105	65	44
1 月 22 日	571	131	444	69	127	62
1 月 23 日	830	259	549	105	281	154
1 月 24 日	1287	444	729	180	558	277
1 月 25 日	1975	688	1052	323	923	365
1 月 26 日	2744	769	1423	371	1321	398
1 月 27 日	4515	1771	2714	1291	1801	480
1 月 28 日	5974	1459	3554	840	2420	619
1 月 29 日	7711	1737	4586	1032	3125	705
1 月 30 日	9692	1982	5806	1220	3886	761
1 月 31 日	11791	2102	7153	1347	4638	752
2 月 1 日	14380	2590	9074	1921	5306	668

2 月 2 日	17205	2829	11177	2103	6028	722
2 月 3 日	20438	3235	13522	2345	6916	888
2 月 4 日	24324	3887	16678	3156	7646	730
2 月 5 日	28018	3694	19665	2987	8353	707
2 月 6 日	31161	3143	22112	2447	9049	696
2 月 7 日	34546	3399	24953	2841	9593	544
2 月 8 日	37198	2656	27100	2147	10098	509
2 月 9 日	40171	3062	29631	2618	10540	444
2 月 10 日	42638	2478	31728	2097	10910	381
2 月 11 日	44653	2015	33366	1638	11287	377
2 月 12 日	46472	1820	34874	1503	11598	312
2 月 13 日	48467	1995	36602	1728	11865	267
2 月 14 日	49970	1503	37884	1282	12086	221

全国情况：这里，N(1)=214 是截至 2020 年 1 月 19 日 24 时，全国累计报告新型冠状病毒感染的肺炎病例数量。N(20)= 34546 是截至 2020 年 2 月 7 日 24 时，全国累计报告新型冠状病毒感染的肺炎病例数量。我们尝试根据这 20 个点拟合出曲线，并根据拟合出的曲线，对接下来 7 天的全国确诊病例数量进行预测。这样，我们的任务就是求解如下最优化问题

此优化问题几个变量尺度差距非常大，是一个高度病态问题，好在维数不高，可以通过牛顿法或高斯牛顿法求解。我用高斯牛顿法求得下图（图 1）。

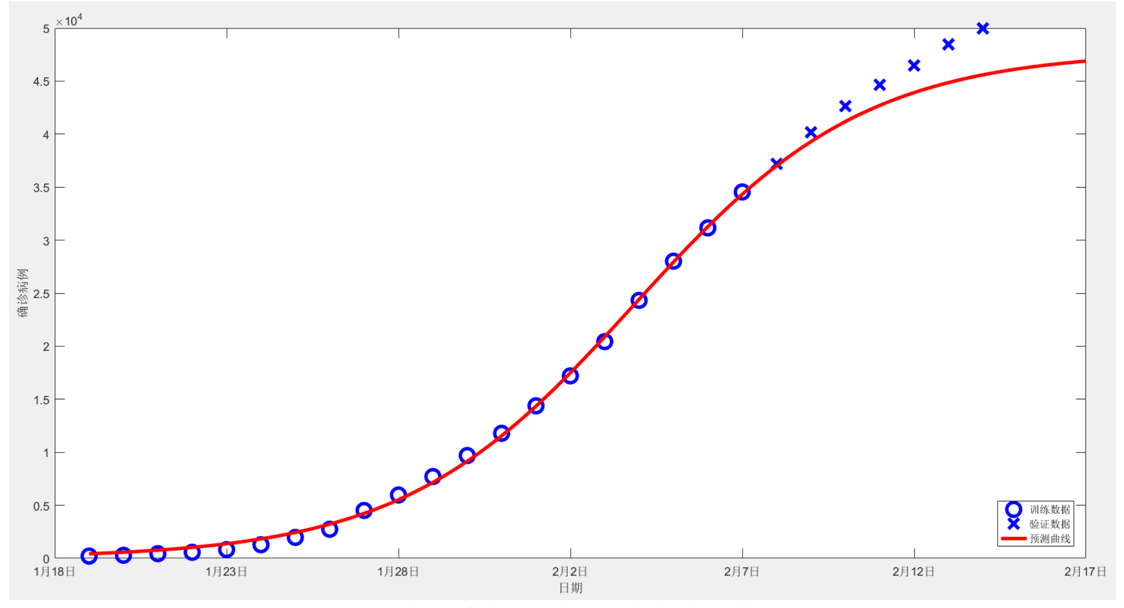


图 1 全国感染新型冠状病毒肺炎情况

在该图中，圈是用来构建模型的训练数据，叉是训练数据后 7 天的真实结果，用于验证预测的可靠性。从图中可以看出，这条曲线的预测并不可靠。它虽然很好地拟合了 2 月 7 日之前的状况，但后面的预测误差较大。

此外在发病时期，湖北省与国内其它地区在医疗条件等方面有明显不同，因此本文将其分开考虑。

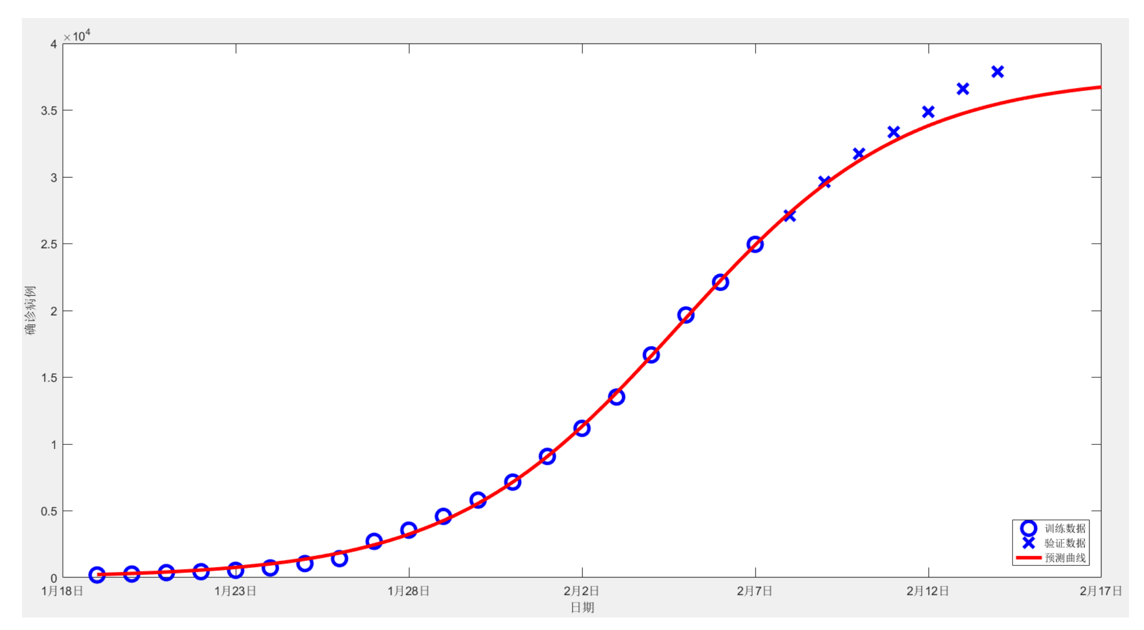


图 2 湖北省感染新型冠状病毒肺炎情况

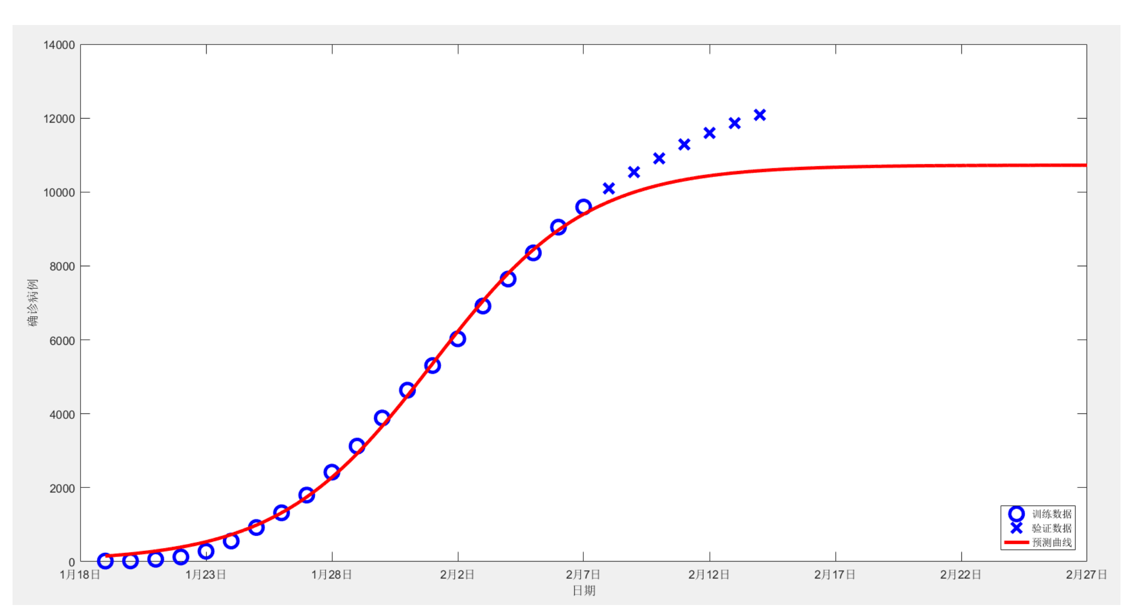


图 3 全国其它地区感染新型冠状病毒肺炎情况

根据上述同样的原理，我们得到图 2 和图 3。尽管它们都很好地拟合了 2 月 7 日之前的数据，但给出的预测与实际有较大差距。

我们猜想，这是由于 1 月 27 日，新的测试试剂开始应用，而这导致确诊率的大幅上升，从而导致确诊人数大幅提升。或者说，可以说，26 日及以前，所有的数据都和真实感染人数有较大的差距。因此，我们又以 1 月 27 日以来的数据重新进行模拟，结果如图 4,图 5 所示。

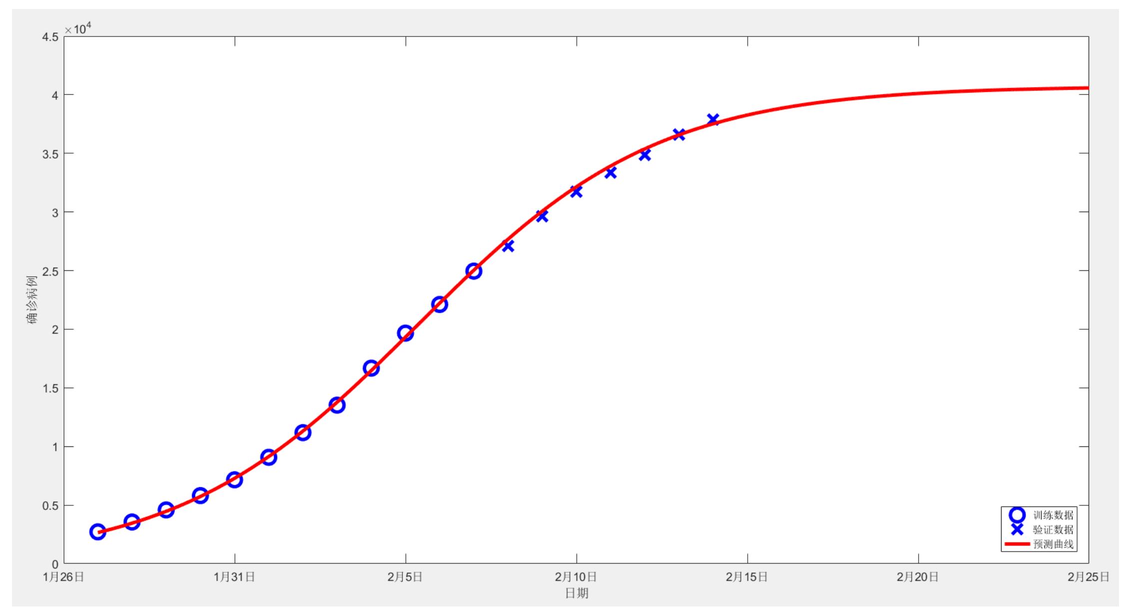


图 4 湖北省感染冠状病毒肺炎情况（1 月 27 日始）

该组数据不仅很好地拟合了 2 月 7 日之前的数据，还对后 7 天的数据提供了很好的估计，拟合效果较好。拟合得到最终湖北省感染新型冠状病毒肺炎人数累计应在 40730 人左右。我们认为，这可能是一个比较保守的估计，真实人数要密切根据动态变化情况予以调整。其实，2 月 5 日以后，每天湖北省确诊人数的增加量都较前日有所降低，可以确认拐点已经过去了。

再考虑全国其它地区的情况，过程与上图类似。模拟估计最终全国除湖北省其他地区感染新型冠状病毒肺炎人数约为 11630 人，但现在感染人数已经超过了这个数字。这条曲线对后面的几天拟合的依旧不好，对此，本文认为可能是由以下几点原因导致的：

1. 模型不够精确。

本文采用了 SI 模型，这个模型在疫情初期是对复杂的 SEIR 模型很好的近似。但是现在疫情已经跨过了拐点很多，并且治愈人数已经有了 8000 多人，这对于总确诊病例近 5 万人来说已经是不可忽略的量了。

2. 确诊率在提高。

随着时间的推移，可能医学研究的发展，医疗物资的状况都有了很大的变化，确诊率一定会随着时间不断提高。

3. 数据不够准确，有误差。

统计数据时由于重复，漏报等因素，会导致确诊病例与真实感染人数有一定差据，这些差距在数据拟合的过程中会放大。

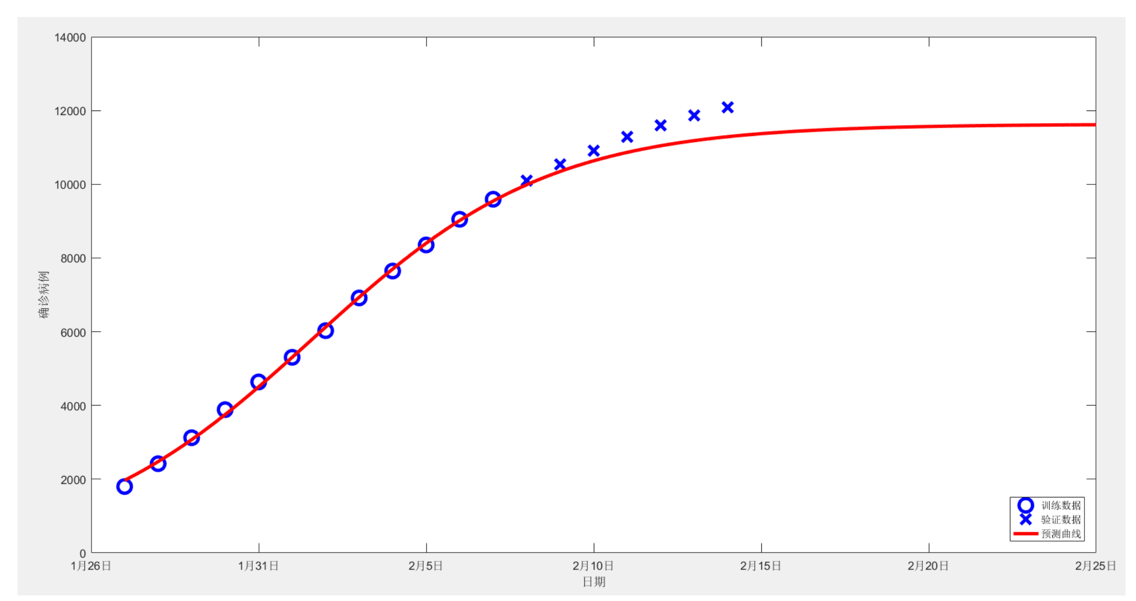


图 5 全国其它地区感染冠状病毒肺炎情况（1 月 27 日始）

用 1 月 27 日到 2 月 7 日的全国感染数据，我们可以得到图 6. 这个预测还是比较可靠的，从图中，我们可以看到，最终全国感染冠状病毒肺炎累计人数大约在 5.3 万人左右。

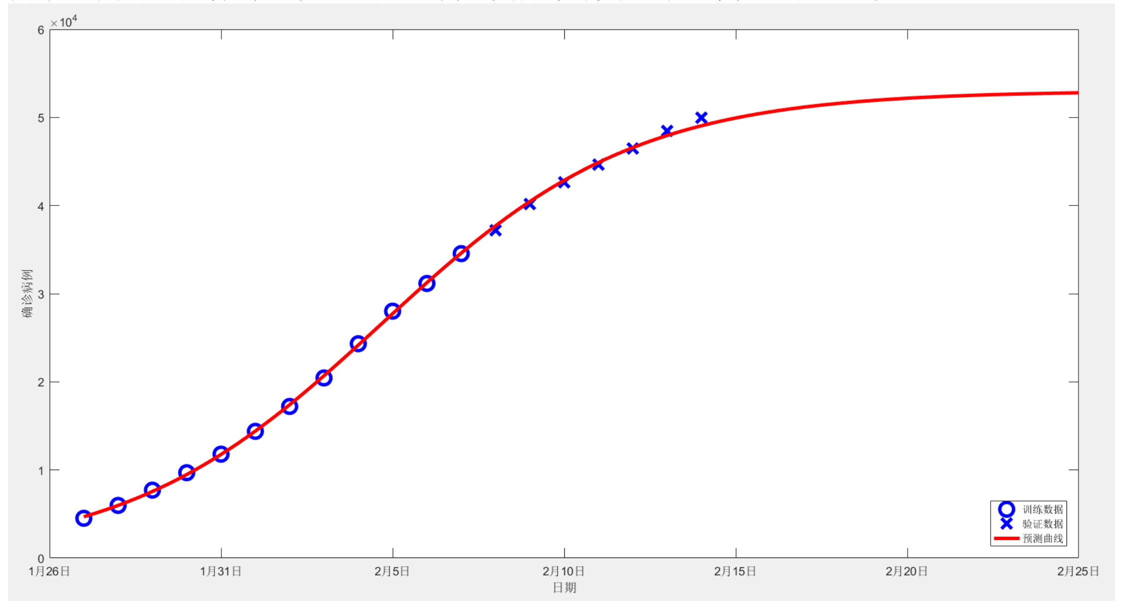


图 6 全国感染冠状病毒肺炎情况（1 月 27 日始）

总之，根据拟合与优化模型计算，湖北省感染新型冠状病毒肺炎的人数累计应在 4.1 万人左右，全国其它地区在 1.3 万人左右，共计约 5.4 万人。未来，可根据数据实时更新拟合结果，以提供较可信的数据。

下面我们讨论本次疫情何时达到拐点的问题。所谓拐点就是新增病例数不再增加。表 1 的实际数据显示，湖北省新增病例在 2 月 4 日达到高峰，但随后数日一直处于波动状态，直至 2 月 9 日以后，才基本呈下降趋势。与湖北省的情况不同，全国其它地区自 2 月 3 日以来，新增病例已经连续下降，应该标明拐点早已经到来。

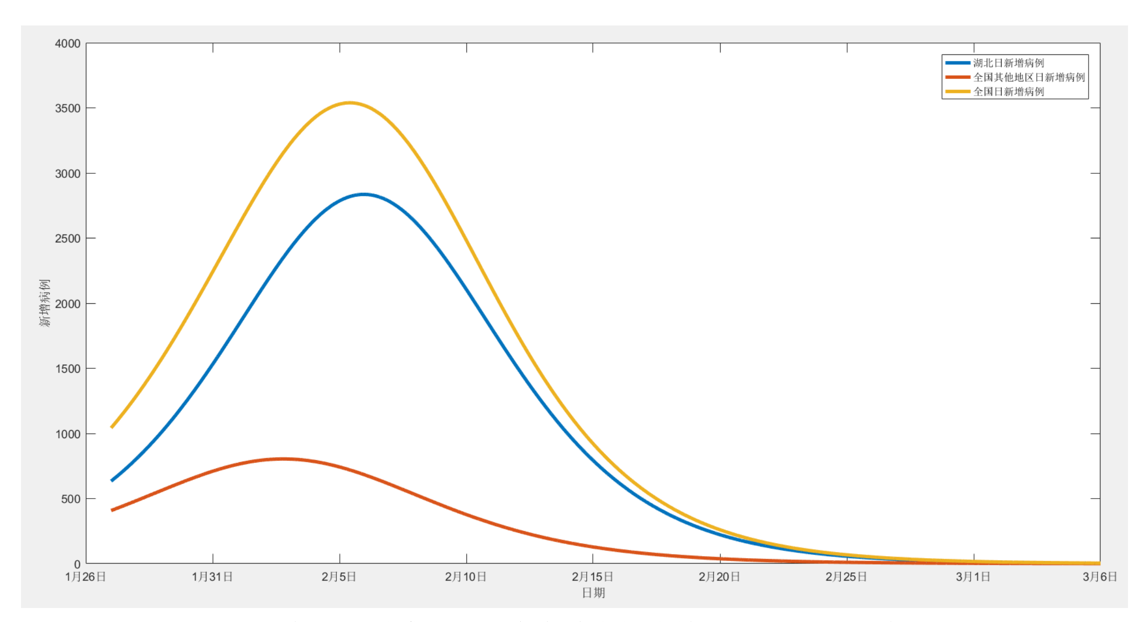


图 7 全国新增感染冠状病毒肺炎拟合情况（1 月 27 日始）

尽管我们的模拟计算表明疫情拐点已过（图 7），但我们仍需提高警惕，防止因防控措施不到位，导致更多的人被传染，从而爆发第二波疫情。

本次疫情何时结束也是大家比较关心的话题。如果上述模拟计算可靠的话，我们以新增病例降至 10 为参考点，湖北省将持续至 3 月 1 日，全国其它地区将持续至 2 月 25 日。也就是说，如果保持目前防控措施的话，本次疫情大约在 3 月初后将得到根本控制。

4 结论与展望

本文采用目前广为通用的 SI 传染病模型，通过拟合和优化计算，获得全国感染冠状病毒肺炎人数大约在 5.4 万人左右，其中湖北省 4.1 万人左右，全国其它地区 1.3 万人左右。

值得指出的，我们的模型对全国其他地区的情况拟合得较差，可能需要采用更复杂的 SEIR 模型，但该模型参数求解难度较大。另一方面，就本文采用的模型而言，也可以尝试加大近期权重，减小较远数据权重的方式解决，这些将在下一步工作中尝试。

2 月 13 日，国家卫健委宣布，新冠肺炎确诊也可以包括临床确诊方法。尽管这一统计口径的变动对全国其它地区的模拟和预测没有任何影响，但将显著提升湖北省感染人数。3 天来，由此而带来的湖北省肺炎感染人数上升 17400 余人，感染总人数达 56000 余人，已大于我们预测的总感染人数。如果这一数据准确可靠的话，湖北省的实际感染人数将超出本文预测数据 40%左右，且疫情结束的日期将推后两周左右，值得相关部门重视。

5 致谢

本文是在导师袁亚湘研究员指导下完成，并得到国家自然科学基金委员会资助（项目编号 11331012、11688101）。两位匿名评审人对论文初稿提出了许多建设性修改意见，使得论文质量得到很大提高，在此一并致谢。同时感谢编辑部葛向阳老师耐心细致的工作。

6 参考文献

[1] Wu J T, Leung K, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak origination in Wuhan, China: a modelling study[J]. The Lancet, 2020, S0140-6736(20)30260-9.

[2]Huang N E, Qiao F L. A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV[J]. Sci. Bull., 2020, 65: doi: 10.1016/j.scib.2020.02.005.

[3]严阅, 陈瑜, 刘可伋, 罗心悦, 许伯熹, 江渝, 程晋. 基于一类时滞动力学系统对新型冠状病毒肺炎疫情的建模和预测[J]. 中国科学: 数学, 2020, 50: doi: 10.1360/SSM-2020-0026.

[4]马知恩, 周义仓. 传染病动力学的数学建模与研究 [M]. 北京: 科学出版社, 2004.

[5] Cantó B, Coll C, Sánchez E. Estimation of parameters in a structured SIR model [J]. Adv. Differ. Equ., 2017, 2017:33.

[6]刘畅, 丁光宏, 龚剑秋, 王凌, 程珂, 张迪. SARS 爆发预测和预警的数学模型研究[J]. 科学通报, 2004, 49(21): 2245-2251.